IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2015), pp. 5625-5629, Brisbane, Australia, April 2015.

ENHANCED ROBOT AUDITION BY DYNAMIC ACOUSTIC SENSING IN MOVING HUMANOIDS

Vladimir Tourbabin[1]*, Hendrik Barfuss*[2]*, Boaz Rafaely*[1]*, and Walter Kellermann*[2]

[1] Dept. of Electr. and Comp. Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel
{tourbabv,br}@ee.bgu.ac.il

[2] Multimedia Comm. and Signal Proc., University of Erlangen-Nürnberg, Erlangen, Germany
{barfuss, wk}@LNT.de

ABSTRACT

Auditory systems of humanoid robots usually acquire the surrounding sound field by means of microphone arrays. These arrays can undergo motion related to the robot's activity. The conventional approach to dealing with this motion is to stop the robot during sound acquisition. This approach avoids changing the positions of the microphones during the acquisition and reduces the robot's ego-noise. However, stopping the robot can interfere with the naturalness of its behaviour. Moreover, the potential performance improvement due to motion of the sound acquiring system can not be attained. This potential is analysed in the current paper. The analysis considers two different types of motion: (i) rotation of the robot's head and (ii) limb gestures. The study presented here combines both theoretical and numerical simulation approaches. The results show that rotation of the head improves the high-frequency performance of the microphone array positioned on the head of the robot. This is complemented by the limb gestures, which improve the low-frequency performance of the array positioned on the torso and limbs of the robot.

Index Terms— Humanoid robots, robot audition, acoustic sensing, DoA estimation, robomorphic array.

1. INTRODUCTION

Auditory systems of humanoid robots is a wide area of research. These systems usually acquire the surrounding sound field by means of microphone arrays. The signal processing methods based on these arrays include sound localization [1], source separation [2], noise suppression [3], echo cancellation [4], and speech recognition [5]. These methods typically assume that the robot and, therefore, the array are fixed in a given position. However, the arrays, installed on the head [6], torso, and limbs [7] of the robot, can move due to rotation of the head, hand gestures, and more complex movements like walking. The conventional approach to motion is largely based on the "stop-perceive-act" principle [3] that suggests stopping the robot during the sound acquisition. This approach has two major drawbacks; first, stopping the robot for sound acquisition may interfere with naturalness of the robot's behaviour, and second, the potential performance improvement from the movement of the sound acquisition system [8] cannot be attained.

In the current paper, we analyse the potential advantages of robot motion on the performance of its auditory system. Note that the effect of ego-noise [9] is left out of the scope of the current paper. The system considered here consists of two different arrays: (i) rigid head array and (ii) robomorphic body array. The analysis is conducted by comparing the performance of the auditory system of a static robot to that of a moving robot. Two types of motion are considered in the current study. The first is rotation of the head, which leads to continuous rotation of the head array. The second is limb gestures, which provide the means for controlling the aperture of the robomorphic array. The two types of arrays are first analysed theoretically for the motion-driven increase in the amount of the acquired information and for their ability to complement each other. This analysis is performed by applying the effective rank measure of array quality [6] as a function of the head rotation velocity and the aperture of the robomorphic array. In addition to the theoretical analysis, an example of the effect of motion on the joint performance of the arrays is provided by evaluating their Direction of Arrival (DoA) estimation accuracy. The emphasis is made on DoA estimation because it is one of the fundamental abilities of a humanoid auditory system; it is used for localization of acoustic events [10, 11], as well as a preprocessing step for Automatic Speech Recognition (ASR) [5].

The paper is structured as follows: Section 2 introduces the measurement model for the head and robomorphic arrays. Sections 3 and 4 present theoretical analysis of the information acquired by the dynamic arrays and a numerical study of their DoA estimation accuracy, respectively. Section 5 concludes the paper.

2. MICROPHONE ARRAY SYSTEMS

This section describes the head and robomorphic arrays and outlines their measurement models. These models will be used in the theoretical analysis and the numerical study that follow.

2.1. Head array

The head array can be described as an array of M_{H} microphones distributed on the surface of the robot's head, as illustrated in Fig. 1(a). The positions of the microphones relative to the head and to each other are fixed. Suppose that the array is rotating about its center in a sound field produced by S fixed sources. Consider the Short-Time Fourier Transform (STFT) [12] of the microphone outputs, $\mathbf{p}_{\mathrm{H}}(\tau, \nu)$, where τ and ν are the time and frequency indices of a given time-frequency (TF) bin, respectively. The STFTs of the microphone outputs can be related to the amplitudes of the sources

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 609465.

by [13]:

$$\mathbf{p}_{\mathrm{H}}(\tau,\nu) = \mathbf{H}(\tau,\nu,\alpha)\mathbf{s}(\tau,\nu) + \mathbf{n}_{\mathrm{H}}(\tau,\nu), \tag{1}$$

where $\mathbf{n}_{\mathrm{H}}(\tau,\nu)$ is an additive noise component, $\mathbf{s}(\tau,\nu) = [s_1(\tau,\nu)\ \dots\ s_S(\tau,\nu)]^T$ holds the source amplitudes with $(\cdot)^T$ denoting the matrix transpose operator. Matrix $\mathbf{H}(\tau,\nu,\alpha)$ describes the propagation of sound from the sources to the microphones with α denoting the rotation velocity in rad/s. The rotation here is limited to counter-clockwise rotation around the z axis without loss of generality. Using the spherical harmonics (SH) domain, this matrix can be expressed as [8]:

$$\mathbf{H}(\tau,\nu,\alpha) = \mathbf{V}(\nu)\mathbf{R}(\alpha T\tau/f_{\mathrm{s}})\mathbf{Y}(\Psi), \tag{2}$$

where $\Psi = \{(\theta_q,\phi_q)\}_{q=1}^{S}$ is the set of the DoAs of all the sources, with θ_q and ϕ_q denoting the elevation and azimuth in the conventional spherical coordinate system [14] positioned at the array center and fixed relative to the sources. Matrix $\mathbf{Y}(\Psi)$ denotes the frequency-independent steering matrix in the SH domain with the columns given by $\mathbf{y}(\theta_q,\phi_q) = [Y_0^0(\theta_q,\phi_q)\, Y_1^{-1}(\theta_q,\phi_q)\ \dots\ Y_N^N(\theta_q,\phi_q)]^H$, where $(\cdot)^H$ denotes the conjugate-transpose operator and $Y_n^m(\cdot,\cdot)$ is the spherical harmonic of order n and degree m. More specifically, the entries of $\mathbf{y}(\theta_q,\phi_q)$ are arranged such that $Y_n^m(\theta_q,\phi_q)$ is the element number n^2+n+m of the vector. The sound field on the surface of the head is assumed to have a limited SH order N. Assuming that the head shape is close to spherical, the effective SH order of the field in frequency bin ν is given by [15] $N = \lceil kr \rceil$, where $k = \frac{2\pi f_{\mathrm{s}}}{Tc}\nu$ is the wave number and r is the sphere radius, with f_{s} denoting the temporal sampling frequency, T is the length of the STFT time frames, and c is the speed of sound. Matrix $\mathbf{R}(\alpha T\tau/f_{\mathrm{s}})$ describes the rotation of the sound field at time frame τ relative to the coordinate system positioned at the array center and moving together with the array. In particular, the product $\mathbf{Y}(\Psi_\tau) = \mathbf{R}(\alpha T\tau/f_{\mathrm{s}})\mathbf{Y}(\Psi)$ is, in fact, the SH steering matrix for the sources during time frame τ as viewed from the coordinate system that is moving together with the array. The entries of $\mathbf{R}(\alpha T\tau/f_{\mathrm{s}})$ are given by the Wigner-D function that can be evaluated using analytic expressions given the rotation angles at time frame τ [8]. Finally, $\mathbf{V}(\nu)$ is given by:

$$\mathbf{V}(\nu) = [\mathbf{v}_{0,0}^*(\nu)\, \mathbf{v}_{1,-1}^*(\nu)\ \cdots\ \mathbf{v}_{N,N}^*(\nu)] \in \mathbb{C}^{M\times(N+1)^2}, \tag{3}$$

where $(\cdot)^*$ denotes the complex-conjugate operator and $\mathbf{v}_{nm}(\nu)$ holds the spherical Fourier transform (SFT) coefficient of the conjugate of the array steering vector of order n and degree m. In practice, the steering vectors can be obtained by measurements [2] or by numerical simulations [6].

2.2. Robomorphic array

The robomorphic array consists of five omnidirectional microphones attached to the robot's torso and limbs, as illustrated in Fig. 1(b), modeling a standing humanoid robot with slightly raised open arms. For both legs and arms, the same microphone spacing d with respect to the center microphone is considered. Moreover, the array is assumed to be located in the x-z plane.

The M_{R} microphone signals of the robomorphic array in the STFT domain are given as:

$$\mathbf{p}_{\mathrm{R}}(\tau,\nu) = \mathbf{A}(\nu,d)\mathbf{s}(\tau,\nu) + \mathbf{n}_{\mathrm{R}}(\tau,\nu). \tag{4}$$

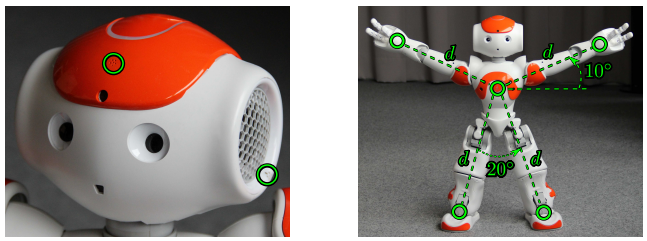
The M_{R}-dimensional vectors $\mathbf{p}_{\mathrm{R}}(\tau,\nu)$ and $\mathbf{n}_{\mathrm{R}}(\tau,\nu)$ contain the microphone signals and additive noise, respectively, whereas the S-dimensional vector $\mathbf{s}$ captures the source signals. Assuming free-field propagation, the $M_{\mathrm{R}} \times S$-dimensional transfer matrix $\mathbf{A}(\nu,d)$ consists of all the steering vectors between the S sources and M_{R} sensors, defined as [13]:

$$\mathbf{a}(\mathbf{k}_q,d) = \left[e^{-j\mathbf{k}_q^T\mathbf{x}_0(d)} \quad e^{-j\mathbf{k}_q^T\mathbf{x}_1(d)} \quad \dots \quad e^{-j\mathbf{k}_q^T\mathbf{x}_{M_{\mathrm{R}}-1}(d)}\right]^T, \tag{5}$$

with wavevector $\mathbf{k}_q$ given as [13]:

$$\mathbf{k}_q = -k\,[\sin(\theta_q)\cos(\phi_q) \quad \sin(\theta_q)\sin(\phi_q) \quad \cos(\theta_q)]^T. \tag{6}$$

Vectors $\mathbf{x}_p(d)$, $p \in \{0,\dots,P-1\}$ in (5) include the position of each microphone in the Cartesian coordinate system, which depends on the microphone spacing d. Note that the use of the free-field steering vectors is a simplifying approximation. A more accurate approximation of steering vectors can be obtained by measurements or numerical simulations of the array in various positions, which is out of the scope of the current study.



(a) Head array. (b) Fixed robomorphic array.

Fig. 1. Illustration of the topologies of the head and robomorphic arrays. Green circles indicate microphone positions.

3. THEORETICAL ANALYSIS

The current section presents a theoretical analysis of the information acquired by the two different arrays described above as a function of frequency and motion parameters. For this purpose, we employ the *effective rank* of the measurement model matrix, which is a theoretical measure of array quality introduced in [6]. The measure was shown to be related to both beamforming and DoA estimation performance.

3.1. Head array

The effective rank measure that we use here suggests to evaluate the array quality by calculating the effective rank $\mathcal{R}(\cdot)$ [16] of the transfer matrix that describes the propagation of sound from the source to the microphones. For example, using the model in (1) for the head array in a given time frame τ, the quality of the array would be evaluated as $\mathcal{R}\left(\mathbf{H}(\tau,\nu,\alpha)\right)$. The measure applies strictly only to stationary arrays. In order to extend the applicability of the measure to moving arrays, recall that a given row in the propagation matrix describes

the propagation to a given microphone. Also note that sampling the sound field by a moving array can be thought of as producing additional virtual microphones at each of the array positions in the different time frames. Motivated by these observations we propose to define a moving-array propagation matrix as a column concatenation of the individual propagation matrices during each time frame, i.e.

$$\mathbf{H}_L(\nu,\alpha) = [\mathbf{H}(0,\nu,\alpha)^T \ \dots \ \mathbf{H}(L-1,\nu,\alpha)^T]^T, \quad (7)$$

where L is the number of adjacent frames that are used. Note that in (7) it is assumed that the two coordinate systems, the one that is moving with the array and the fixed system, are aligned for $\tau = 0$.

The effective rank of $\mathbf{H}_L(\nu,\alpha)$ for a 2-microphone array on the head of the humanoid robot Nao [17] was evaluated. The microphones are positioned on the top front and on the bottom left of the head (see Fig. 1(a)). These positions form a subset of the existing (Phase I) 4-microphone array on the head of this robot. The average radius of the head is 6.25 cm. The effective rank of $\mathbf{H}_L(\nu,\alpha)$ associated with this array was evaluated as a function of frequency ν and for different rotation velocities α. The number of time frames was $L = 40$. This value was chosen in order that the DoA algorithms applied in the next section for the evaluation of performance will produce an estimate each 0.5 seconds. The steering vectors were calculated numerically using the boundary element method (BEM) [6] and the geometry of the robot's head surface. The results are plotted in Fig. 2(a).

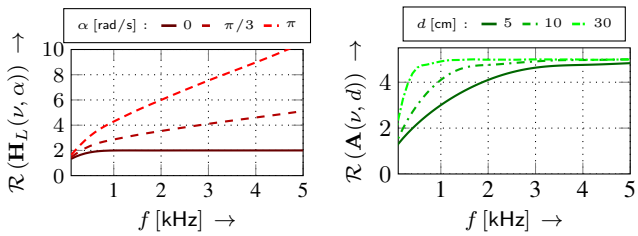
It can be seen that, regardless of the rotation velocity, the effective rank increases with frequency. This is due to the increasing ratio between the spatial sensor separation and the wavelength, which, in turn, increases the phase related to propagation between the sensors, thereby making the array more sensitive to source direction. It is important to note that the maximum effective rank of the stationary array achieved at higher frequencies is 2. This equals the number of microphones in the array, which is also the number of rows in the array propagation matrix. The effective rank of the rotating array is not limited to the number of microphones and significantly outperforms the stationary array at higher frequencies.

3.2. Robomorphic array

In this subsection, the effective rank of the fixed robomorphic array (meaning no movement of the limbs is exploited) is analysed. Fig. 2(b) illustrates the effective rank of $\mathbf{A}(\nu,d)$ in (4) for different microphone spacings $d \in \{5,10,30\}$ cm. It can be seen that the larger the microphone spacing is, the faster the effective rank of the matrix increases, reaching full rank at $f \approx 1500$ Hz for $d = 30$ cm.

3.3. System complementarity

The effective ranks of the head and robomorphic arrays are compared in Fig. 3. The comparison focuses on low frequencies to facilitate the analysis of individual array advantages. The robomorphic array has a larger aperture, compared to the head array. This fact provides the robomorphic array with an advantage up to $f = 1500$ Hz. On the other hand, above 1500 Hz, the effective rank of the rotating head array is higher. These results imply that the two arrays can complement each other in different frequency ranges and the performance of the combined array is expected to be higher than the performance of each array alone. Note also that it is possible that the effective rank of the robomorphic array can be further increased by utilizing movements of the robot's limbs and body. However, this is out of the scope of the current paper and has not been investigated here.



(a) Rotating head array. (b) Fixed robomorphic array.

Fig. 2. Effective rank of the rotating head array and of the robomorphic array as a function of frequency for different rotation velocities α and microphone spacings d, respectively.

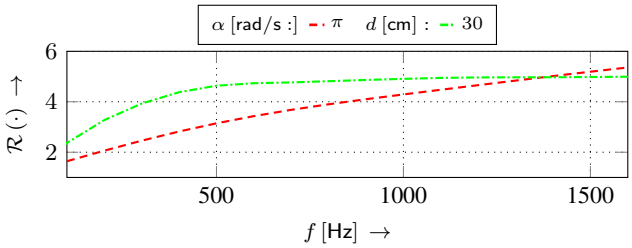


Fig. 3. Comparison of the effective ranks of the 2-microphone rotating head array and the 5-microphone fixed robomorphic array as a function of frequency.

4. DOA ESTIMATION PERFORMANCE

In this section, we study the DoA estimation performance of both the rotating head array and the fixed robomorphic array. The scenario considered here consists of a single source in free field. Spatially white Gaussian noise with a power corresponding to a wide-band Signal-to-Noise Ratio (SNR) of 10 dB was added to the outputs of the microphones. For both arrays, $L = 40$ consecutive snapshots were used. The DoA estimation was carried out using the following 9 source directions: $(\theta,\phi) \in \{(90°,90°),(76°,55°),(104°,125°),(57°,107°),(123°,73°),(90°,60°),(90°,120°),(60°,90°),(120°,90°)\}$. These positions were chosen in order to imitate interaction of the robot with a frontal source. For each source position, ten different realizations of white Gaussian noise where used as the source signals. The white noise was chosen in order to avoid frequency-dependent SNR. This way, it is possible to study the array performance over a broad frequency range, which may cover different sources including low frequency noise and speech. For the estimation, blocks with length $T = 256$ samples and overlap of 50%, were used. The sampling frequency was $f_s = 10$ kHz. The localization performance was assessed through the standard deviation of the angle δ between the estimated and true arrival directions. Furthermore, no outliers have been excluded before calculating the standard deviation. The results presented in the following sections were obtained by taking the average of the individual results for all source positions and source signals.

4.1. Head array

The outputs of the microphones in the rotating array described in the previous section were simulated using the overlap-save technique with the filters changing each millisecond. The filters for each required direction were formed using the steering vectors calculated numerically, as mentioned in the previous section. The estimation of the DoA was based on the model in (1) and utilized the Space-Domain Distance (SDD) algorithm introduced in [12]. This is not a subspace algorithm like MUltiple SIgnal Classification (MUSIC). It is used here because of its ability to utilize the array motion by combining measurements taken at different array positions in a way similar to the Synthetic-Aperture Radar (SAR) technique [18]. The SDD algorithm estimates the angles $(\hat{\theta},\, \hat{\phi})$ in each subband ν as follows:

$$(\hat{\theta},\, \hat{\phi}) = \underset{(\theta,\, \phi)}{\operatorname{argmax}} \left(\sum_{\tau=0}^{L-1} d\left(\mathbf{p}_{\mathrm{H}}(\tau,\nu), \mathbf{H}(\tau,\nu,\alpha)\mathbf{y}(\theta,\phi)\right) \right)^{-1}, \tag{8}$$

where $d(\cdot,\cdot)$ measures the tangent between two vectors. For additional details of this algorithm, the reader is referred to [12]. The DoA estimation performance of the array was evaluated as a function of frequency and rotation velocity. The results are presented in Fig. 4.

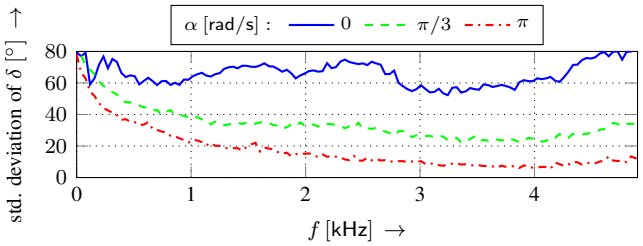


Fig. 4. DoA estimation performance of the rotating head array as a function of frequency for various rotation velocities α.

It can be seen that the stationary array is incapable of DoA estimation with reasonable accuracy over the whole frequency range. This is because the array contains only 2 microphones and so is incapable of DoA estimation in three dimensions. However, the performance improves when rotation is employed. The improvement is more significant at higher frequencies, which is in agreement with the analysis of the effective rank in the previous section.

4.2. Robomorphic array

For the robomorphic array, the spectral MUSIC algorithm [13, 19] is employed for DoA estimation. The estimated angles $(\hat{\theta},\, \hat{\phi})$ in each subband ν are given as:

$$(\hat{\theta},\, \hat{\phi}) = \underset{(\theta,\, \phi)}{\operatorname{argmax}} \left(\mathbf{a}^H(\theta,\phi)\hat{\mathbf{U}}_{\mathrm{N}}\hat{\mathbf{U}}_{\mathrm{N}}^H\mathbf{a}(\theta,\phi) \right)^{-1}, \tag{9}$$

where $\mathbf{a}^H(\theta,\phi)$ is the array steering vector according to (5) and $\hat{\mathbf{U}}_{\mathrm{N}}$ consists of the $M_{\mathrm{R}} - S = M_{\mathrm{R}} - 1$ eigenvectors corresponding to the $M_{\mathrm{R}} - 1$ smallest eigenvalues of the spectral correlation matrix $\hat{\mathbf{R}}_{\mathbf{xx}}(\nu) = 1/L \sum_{\tau=1}^{L} \mathbf{x}(\tau,\nu)\mathbf{x}^H(\tau,\nu)$, where L denotes the number of STFT blocks, as before. The relative delays of the source signals required for simulating the microphone outputs were chosen according to the DoAs of the sources. In Fig. 5, the DoA estimation performance of the fixed robomorphic array as a function of frequency for various microphone spacings $d \in \{5, 10, 30\}$ cm is illustrated. It can be seen that the performance depends on the array aperture. Increasing the aperture improves the estimation accuracy at low frequencies, as is expected from the above analysis of the effective rank. However, the larger the aperture the lower the frequency at which the performance degrades, due to spatial aliasing. The effect of aliasing is complex; it depends on array aperture, topology, and source positions. This makes the high-frequency performance unpredictable.

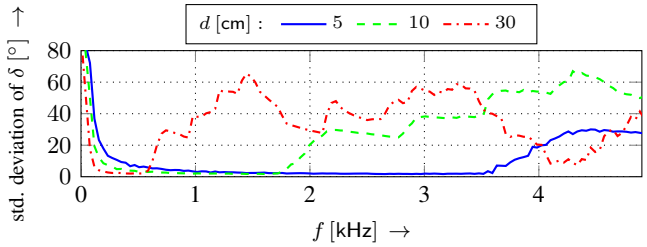


Fig. 5. DoA estimation performance of the robomorphic array as a function of frequency for various microphone spacings d.

4.3. System complementarity

As can be seen from Fig. 5, the high-performance frequency range of the robomorphic array shifts to higher frequencies when the aperture is decreased. However, the aperture can not be lowered indefinitely; it is limited from below by the ability of the robot to shrink its limbs. The operating frequency range can be extended by complementing the robomorphic array with the rotating head array. In particular, by using both arrays simultaneously, high estimation accuracy can be obtained over the whole frequency range. This can be achieved by increasing the aperture of the robomorphic array, thereby focusing it on low frequencies, while, at the same time, the head array will provide high estimation accuracy at high frequencies.

5. CONCLUSION

In the current paper, a dynamic sensing approach to humanoid robot audition was explored by considering a rotating head and controllable limbs. Rotation of the robot's head results in rotation of the head microphone array, whereas the the control over movements of the limbs provides the robomorphic array with a controllable aperture size. Theoretical analysis of the arrays using the effective rank measure showed that the rotation of the head is expected to improve the array performance at high frequencies, while the control over the aperture of the robomorphic array can improve the performance at low frequencies even if this is fixed for the given observation interval. These findings were supported by a numerical study of the DoA estimation accuracy. The results demonstrate that the performance can be improved when motion of the robot is utilized and that the two different arrays can complement each other at different frequencies. Future work may focus on exploring the potential of using both arrays in motion and on the development of localization and signal extraction algorithms that exploit all sensors simultaneously.

6. REFERENCES

[1] U. Kim and H. G. Okuno, "Improved binaural sound localization and tracking for unknown time-varying number of speakers," *Advanced Robotics*, vol. 27, no. 15, pp. 1161–1173, July 2013.

[2] M. Maazaoui, K. Abed-Meraim, and Y. Grenier, "Adaptive blind source separation with HRTFs beamforming preprocessing," in *Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2012, pp. 269–272.

[3] K. Nakadai, T. Lourens, H. G. Okuno, and H. Kitano, "Active audition for humanoid," in *17th National Conference on Artificial Intelligence (AAAI)*, August 2000, pp. 832–839.

[4] R. Takeda, K. Nakadai, T. Takahashi, K. Komatani, T. Ogata, and H.G. Okuno, "Ica-based efficient blind dereverberation and echo cancellation method for barge-in-able robot audition," in *IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, 4 2009, pp. 3677–3680.

[5] K. Nakamura, K. Nakadai, and H. G. Okuno, "A real-time super-resolution robot audition system that improves the robustness of simultaneous speech recognition," *Advanced Robotics*, vol. 27, no. 12, pp. 933–945, May 2013.

[6] V. Tourbabin and B. Rafaely, "Theoretical framework for the optimization of microphone array configuration for humanoid robot audition," *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, vol. 22, no. 12, pp. 1803–1814, December 2014.

[7] H. Barfuss and W. Kellermann, "An adaptive microphone array topology for target signal extraction with humanoid robots," in *Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, September 2014.

[8] V. Tourbabin and B. Rafaely, "Utilizing motion in humanoid robots to enhance spatial information recorded by microphone arrays," in *Joint Workshop Hands-free Speech Communication and Microphone Arrays (HSCMA)*, May 2014, pp. 147–151.

[9] T. Tezuka, T. Yoshida, and K. Nakadai, "Ego-motion noise suppression for robots based on semi-blind infinite non-negative matrix factorization," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, May 2014, pp. 6293–6298.

[10] J.M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2003. (IROS 2003)*, Oct 2003, vol. 2, pp. 1228–1233.

[11] K. Nakamura, K. Nakadai, and G. Ince, "Real-time super-resolution sound source localization for robots," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2012, pp. 694–699.

[12] V. Tourbabin and B. Rafaely, "Speaker localization by humanoid robots in reverberant environments," in *IEEE 28th Convention of Electrical Electronics Engineers in Israel (IEEEI)*, Dec. 2014, pp. 1–5.

[13] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, Detection, Estimation, and Modulation Theory. Wiley, 2004.

[14] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, Elsevier, 2005.

[15] B. Rafaely, "Analysis and design of spherical microphone arrays," *Speech and Audio Processing, IEEE Transactions on*, vol. 13, no. 1, pp. 135–143, January 2005.

[16] O. Roy and M. Vetterli, "The effective rank: A measure of effective dimensionality," in *European Signal Processing conference (EUSIPCO)*, September 2007, pp. 606–610.

[17] Aldebaran Robotics, *NAO NEXT Gen H25 Datasheet*, December 2011.

[18] N. Yen and W. Carey, "Application of syntheticaperture processing to towedarray data," *J. Acoust. Soc. Am.*, vol. 86, no. 2, pp. 754–765, 1989.

[19] R.O. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on*, vol. 34, no. 3, pp. 276–280, March 1986.